History Matching and Emulation for Complex Computer Models

Andrew Iskauskas

Durham University

November 10, 2022

Table of Contents

Complex Models of Real-World Phenomena

Complex computer models (or *simulators*) are used in a variety of fields, including

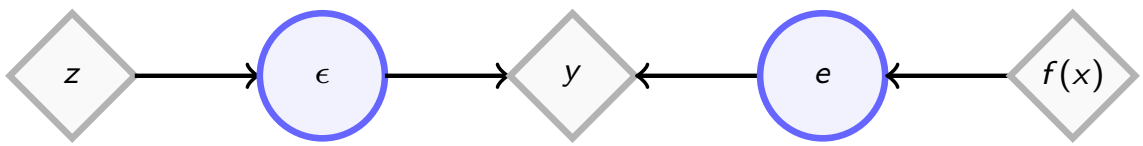
- Oil Industry (oil reservoir and geology models)
- Climate Science (climate models of global warming)
- Systems Biology (genetic and metabolic network models)
- Cosmology (galaxy formation simulations)
- Nuclear Physics (quantum many-body models of nuclei)

Any such simulator is an imperfect or incomplete representation of the reality it seeks to model.

Uncertainty Structure for Models

Consider a simulator $f(x)$ that represents a physical process y, from which we may obtain observed quantities z. Two main sources of uncertainty are

- **Observational error**. Our observations z of y are made imperfectly: $z = y + \epsilon$;
- **Model discrepancy**. Our simulator $f(x)$ cannot faithfully represent the process y: $y = f(x) + e$.



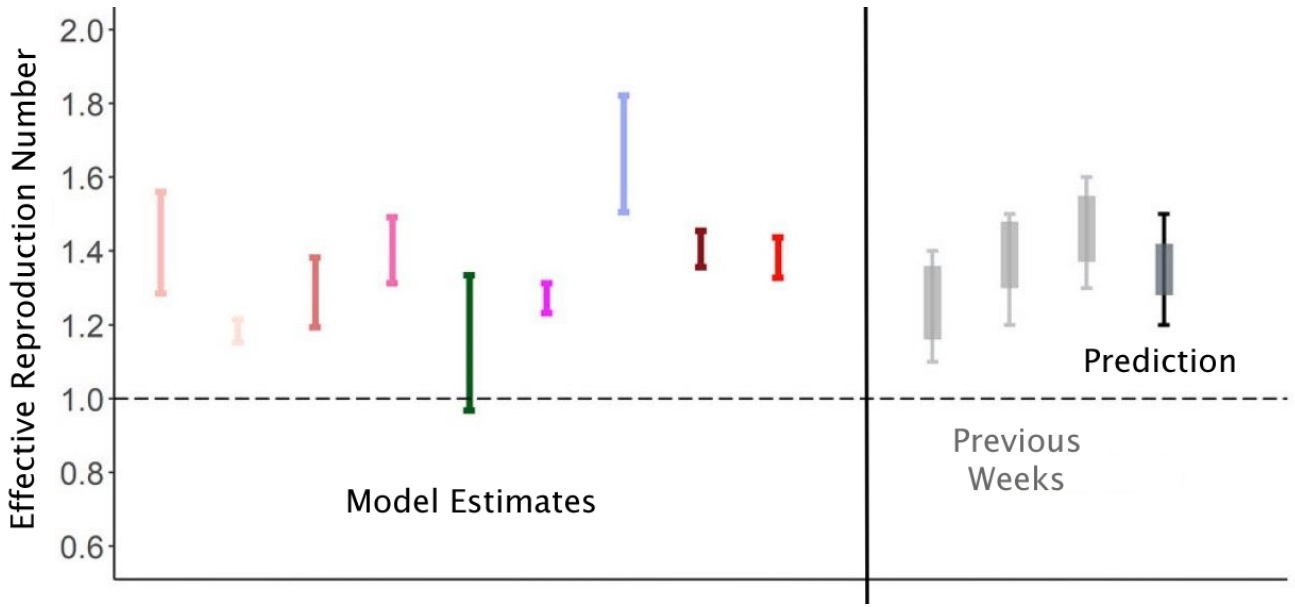
UQ in Epidemiology: I

Simulations are commonly used to represent characteristics of a disease, and finding combinations of parameters that can give rise to observed characteristics of a disease is an important goal.

Without an understanding of the uncertainties that underlie a model, there is always a risk of making incorrect or over-confident predictions.

UQ in Epidemiology: II

COVID-19: Multiple models, with different assumptions, can provide completely different predictions! But they are all modelling the same disease. [1]



[1] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/931146/S0801_SAGE61_201007_SPI-M-O_Consensus_Statement.pdf

The Emulator

An *emulator* is a statistical approximation of a complex computer simulator.

Let $f(x)$ be an output from the simulator at a given parameter set $x \in \mathbb{R}^d$, corresponding to some real physical process y. Then we define a emulator for output $f(x)$ as

$$g(x) = \sum_i \beta_i h_i(x) + u(x).$$

The $h_i(x)$ are a collection of basis functions, β_i their coefficients, and $u(x)$ is a weakly stationary second-order process.

The Emulator

An *emulator* is a statistical approximation of a complex computer simulator.

Let $f(x)$ be an output from the simulator at a given parameter set $x \in \mathbb{R}^d$, corresponding to some real physical process y. Then we define a emulator for output $f(x)$ as

$$g(x) = \sum_i \beta_i h_i(x_A) + u(x_A) + w(x)$$

The $h_i(x_A)$ are a collection of basis functions in the *active variables* x_A, β_i the coefficients, $u(x_A)$ a weakly stationary process in the active variables, and $w(x)$ a 'nugget term'.

Bayes Linear Emulators

Need to provide specifications for the random quantities β, $u(x)$, $w(x)$, then update in light of data. We may be unwilling or unable to provide a full distributional specification.

Instead we take a *pragmatic* approach and require only the specification of second-order quantities: expectations, variances, and covariances. This is the *Bayes Linear* approach.

The Bayes Linear Update Equations

Let $D = \{f(x_1), f(x_2), \ldots, f(x_n)\}$ be runs from the simulator at points $x_1, \ldots, x_n$. The Bayes linear update equations give the emulator's posterior prediction for the model output at an unseen point x, given D:

$$\mathrm{E}_D[g(x)] = \mathrm{E}[g(x)] + \mathrm{Cov}[g(x), D]\mathrm{Var}[D]^{-1}(D - \mathrm{E}[D]),$$
$$\mathrm{Var}_D[g(x)] = \mathrm{Var}[g(x)] - \mathrm{Cov}[g(x), D]\mathrm{Var}[D]^{-1}\mathrm{Cov}[D, g(x)].$$

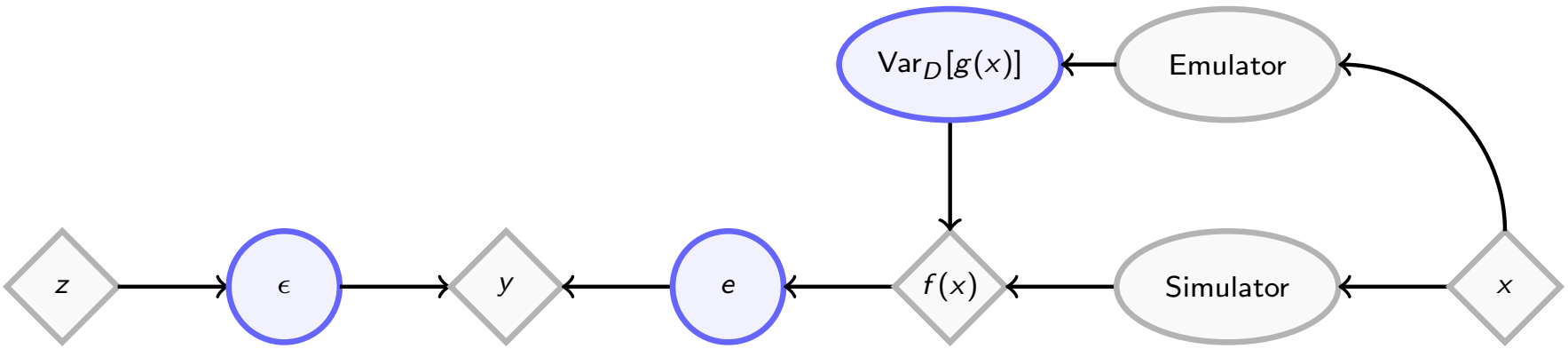
Emulation

Emulators are **fast to evaluate**, requiring only matrix multiplication. For complex models which can take anywhere from minutes to months to evaluate a limited ensemble of runs, an emulator can quickly investigate model behaviour across the entire parameter space.

Emulators **have uncertainty statements built-in**. Each prediction comes with a corresponding uncertainty, $\mathrm{Var}_D[g(x)]$, which depends on the data provided to it and the proximity of the unseen points thereof.

Uncertainty Structure: Emulation

We can extend our chain of uncertainties to link emulator predictions to observations of reality.



Example: 1d Model

Consider a model which can be described in closed form as

$$f(x) = 2x + 3x \sin\left(\frac{5\pi(x - 0.1)}{0.4}\right)$$

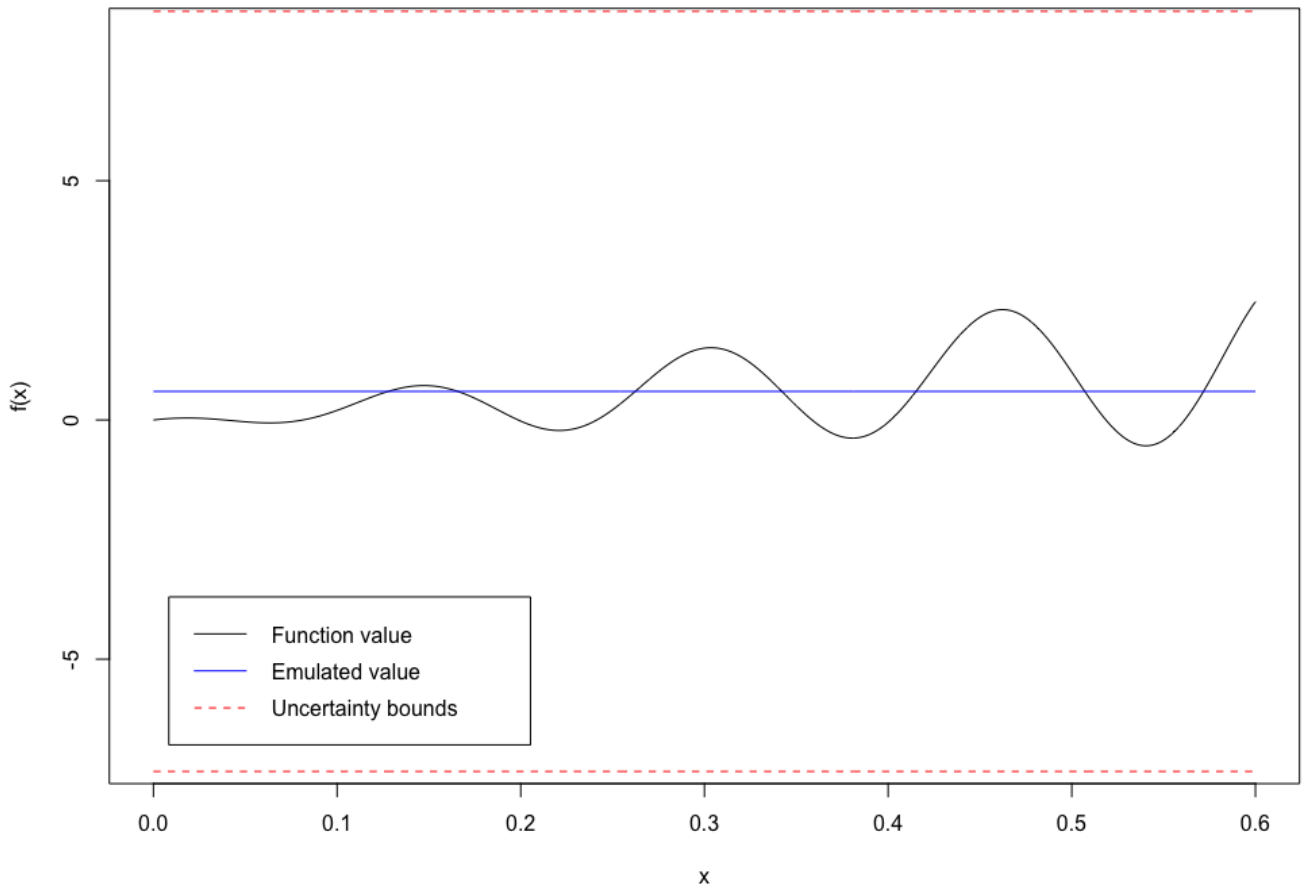
Prior specifications: assume a constant regression function $h(x) = 1$, $\mathrm{E}[\beta] = 0.6$, $\mathrm{Var}[\beta] = 0$, $\mathrm{E}[u(x)] = 0$, and

$$\mathrm{Cov}[u(x), u(x')] = \sigma^2 \exp\left\{-\frac{(x - x')^2}{\theta^2}\right\},$$

with $\sigma^2 = 7$ and $\theta = 0.3$.

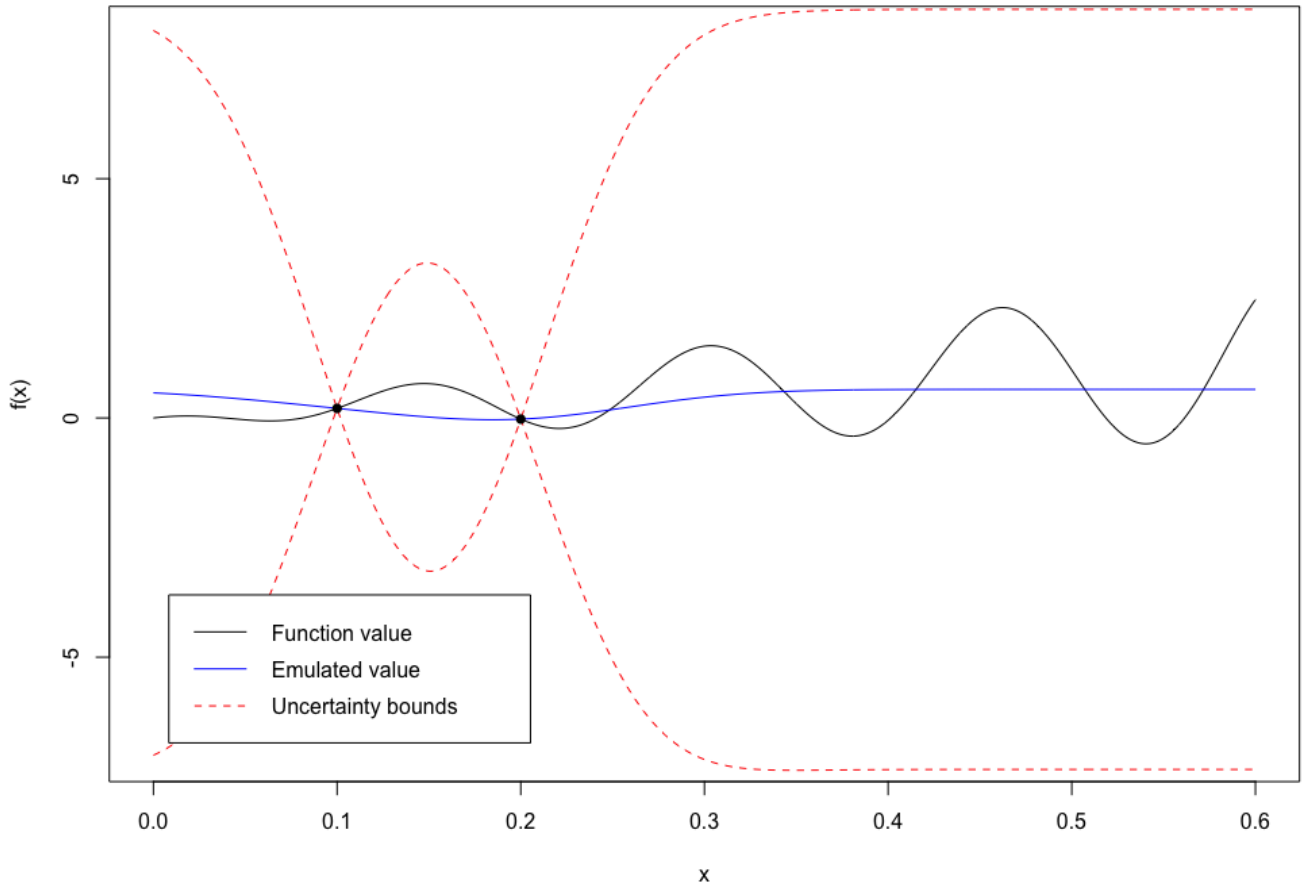
Example: 1d Model

Adjust this prior structure with reference to a collection of 10 points, spaced equally over the interval $[0, 0.5]$.



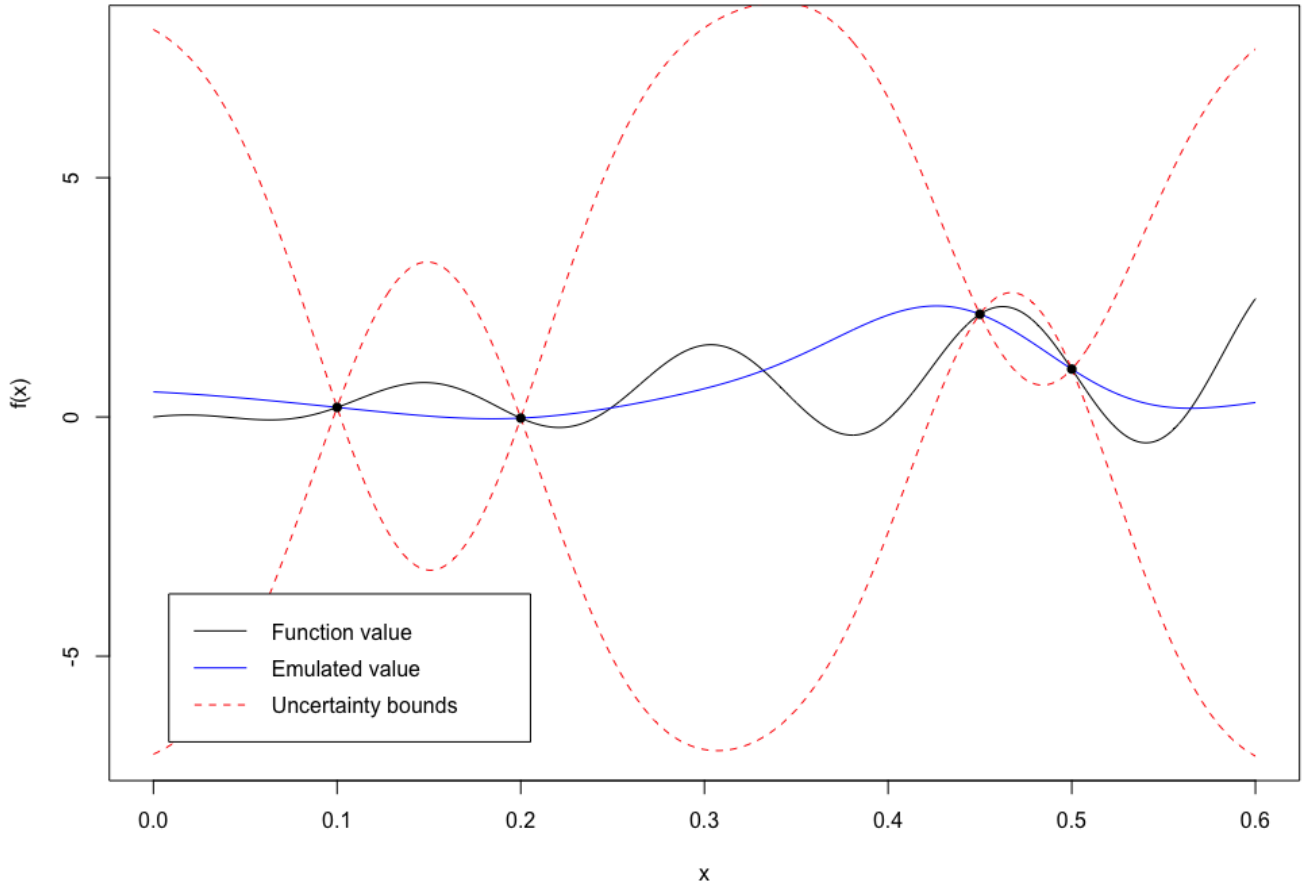
Example: 1d Model

Adjust this prior structure with reference to a collection of 10 points, spaced equally over the interval $[0, 0.5]$.



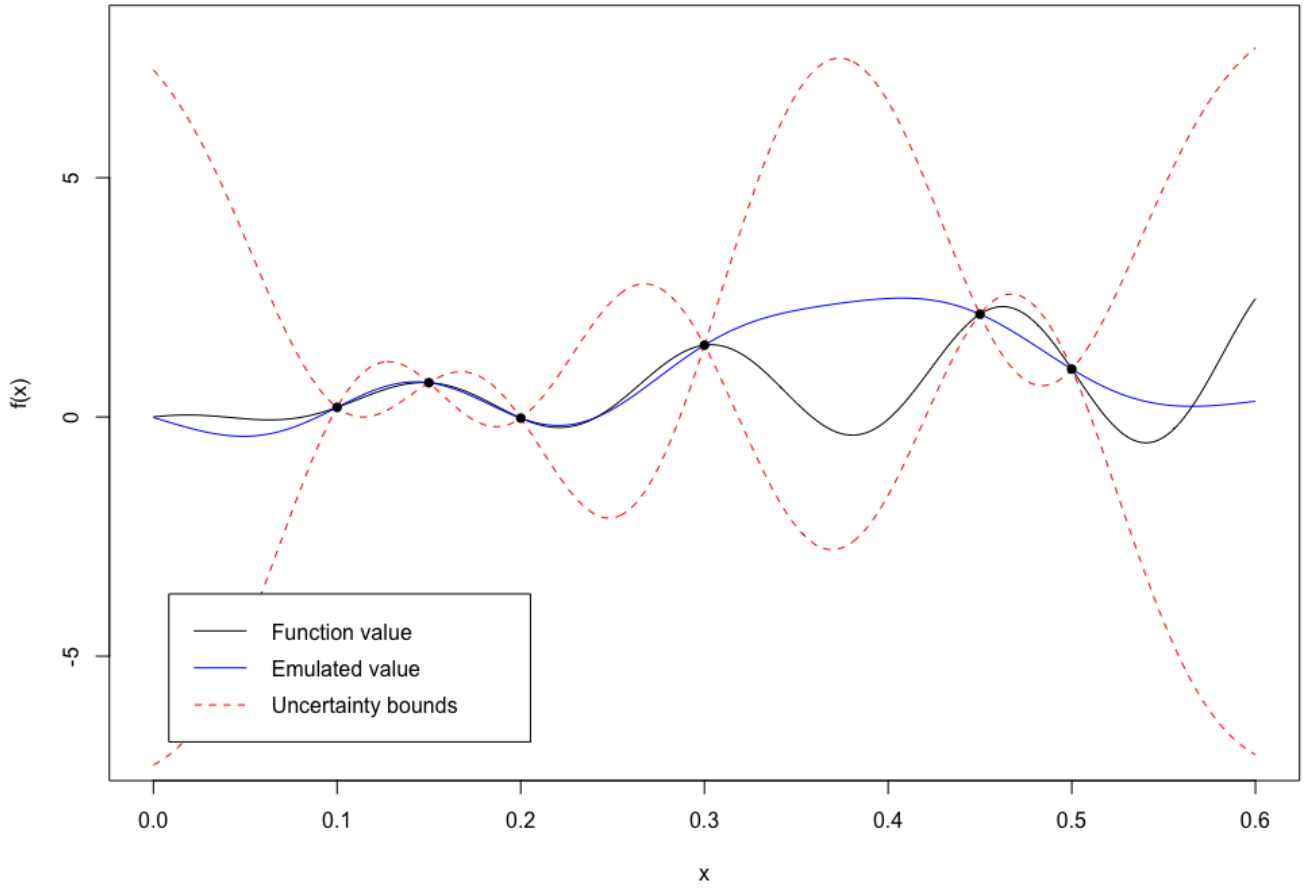
Example: 1d Model

Adjust this prior structure with reference to a collection of 10 points, spaced equally over the interval $[0, 0.5]$.



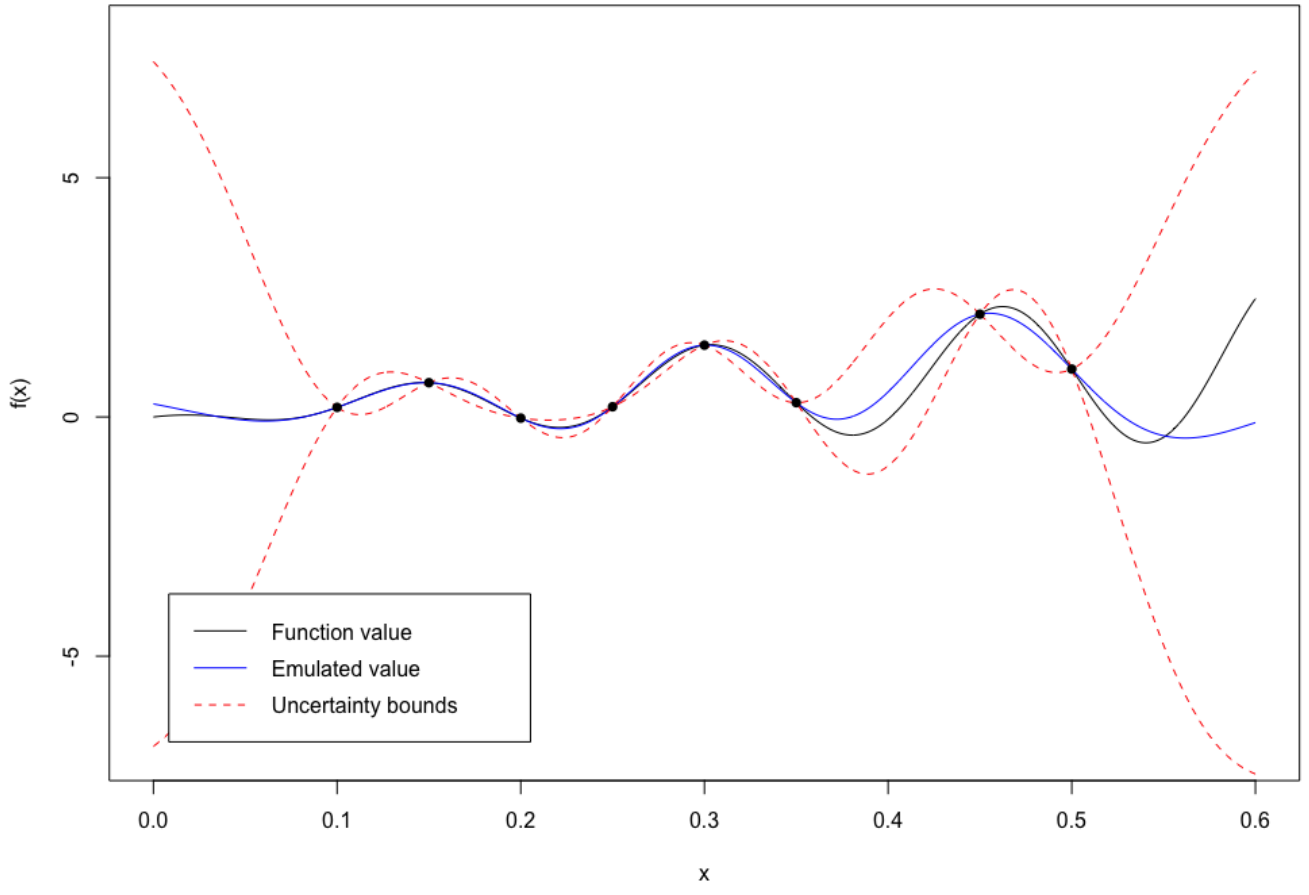
Example: 1d Model

Adjust this prior structure with reference to a collection of 10 points, spaced equally over the interval $[0, 0.5]$.



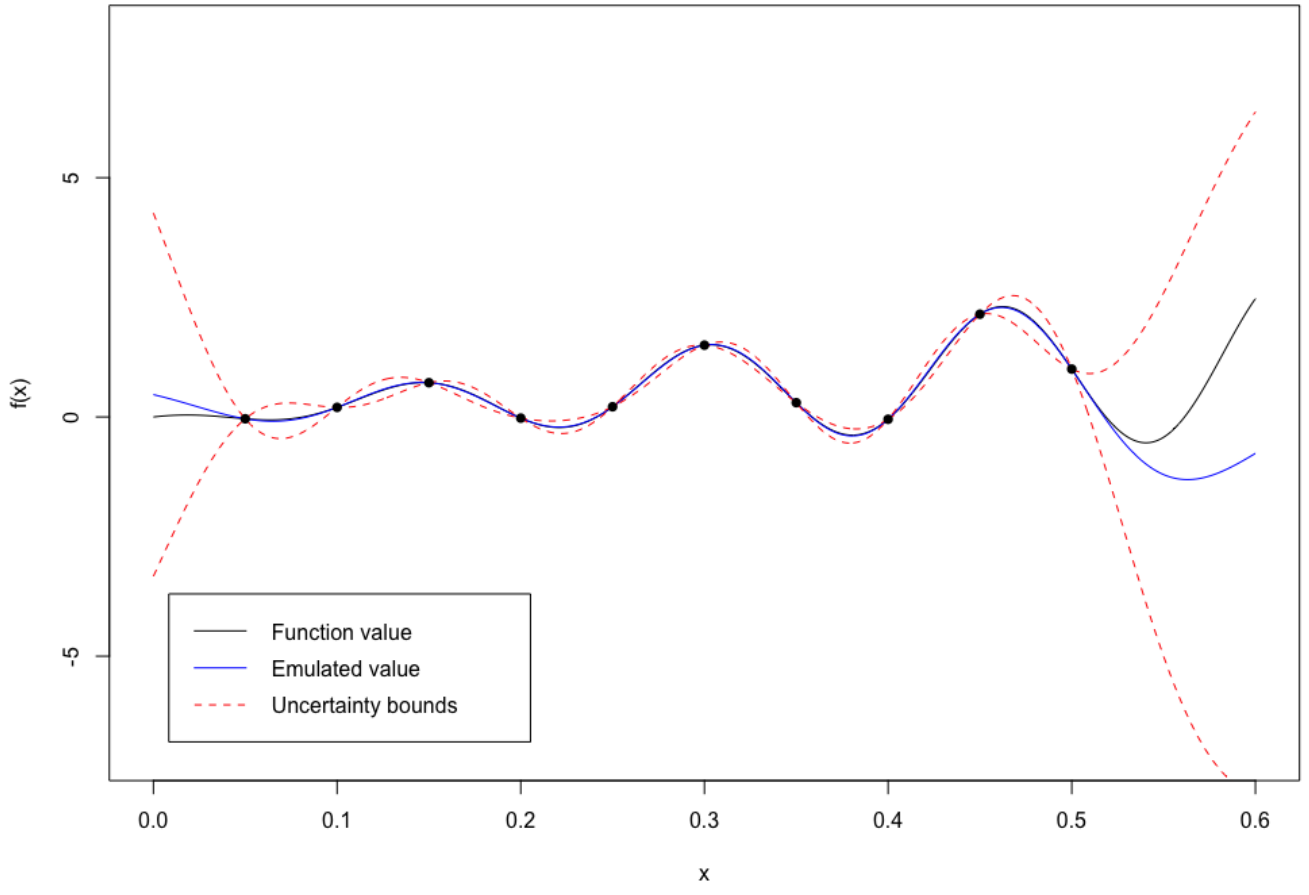
Example: 1d Model

Adjust this prior structure with reference to a collection of 10 points, spaced equally over the interval $[0, 0.5]$.



Example: 1d Model

Adjust this prior structure with reference to a collection of 10 points, spaced equally over the interval $[0, 0.5]$.



History Matching

We have a means of efficiently predicting the output of a complex simulator, given a small sample of runs from it. We want to couple this to a method of *calibration*. The problem statement is:

> Given observed data corresponding to a simulator output, what combinations of input parameters could give rise to output consistent with this observation?

The history matching approach allows this question to be addressed, as well as providing a framework for robustly classifying the various sources of uncertainty.

Implausibility

Before defining the history matching procedure, we need a measure of 'suitability' of an input parameter set x with respect to an observed value z. Define the *implausibility*

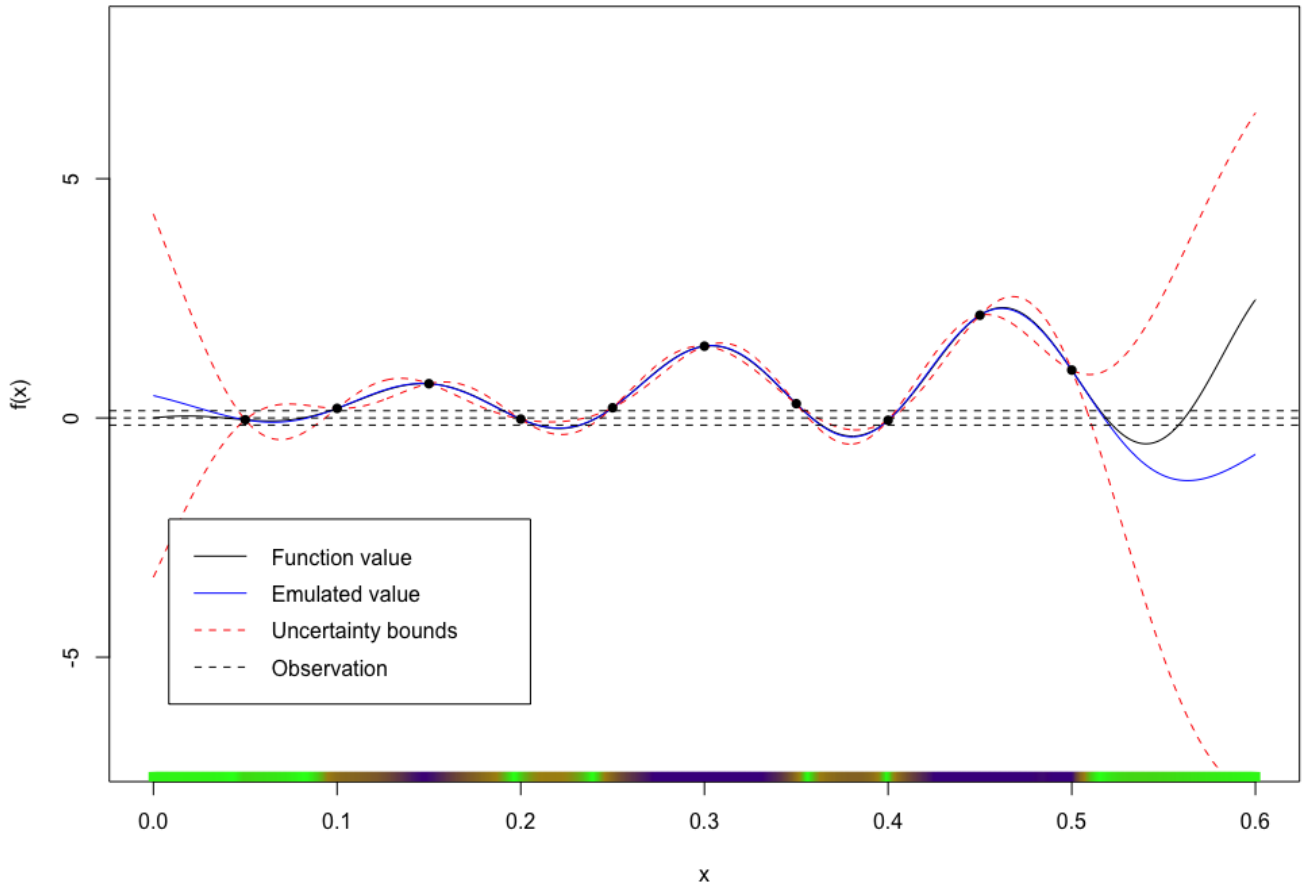
$$I^2(x) = \frac{(\mathrm{E}_D[g(x)] - z)^2}{\mathrm{Var}_D[g(x)] + \mathrm{Var}[e] + \mathrm{Var}[\epsilon]}.$$

The implausibility of a point can be low for two reasons:

- The emulator prediction is close to the observed value, suggesting a good fit;
- The emulator uncertainty at x is large, suggesting a part of parameter space that bears further investigation.

Implausibility: Example

Applied to the one-dimensional example: take an observation of y to be $z = 0$, with uncertainty such that $\mathrm{Var}[\epsilon] = 0.0025$.



The History Matching Principle

Principle of complementarity: a point x is considered unsuitable if **even accounting for the uncertainties in the system**, the prediction $\mathrm{E}_D[g(x)]$ cannot be close to the observed value z.

A point with low implausibility is termed *non-implausible* or *not-yet-ruled-out*.

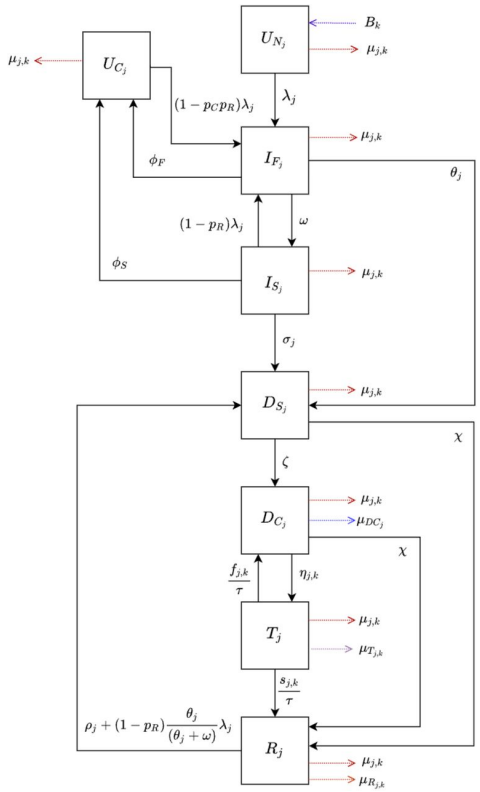
Emulation and history matching work iteratively; each *wave* consists of training emulators on simulator points and proposing non-implausible points from these to run in the simulator.

Can choose to match to a subset of observations at a wave, with no consequences for inferential validity.

The TBvax model

- Deterministic hybrid model of tuberculosis and HIV, designed by colleagues at London School of Hygiene and Tropical Medicine
- Between 19 and 34 input parameters, and 9 to 30 outputs per country
- Run-time for a parameter set varied between seconds and minutes
- Modelling requested by the World Health Organisation in order to form the basis for policy decisions
- Expected to be used to match to observational data for 115 countries

TBvax structure



- Complex compartmental model, even in the absence of HIV progression
- Transition through multiple compartments: uninfected, infected, diseased, treatment, recovery
- Compartments further split by, e.g., clinical and subclinical disease
- Matching to incidence, notifications, deaths, aggregated by age group and year

Challenges

- The number of countries required, and the timescale provided, was such that a supervised process of calibration for each country was impossible.
- TBvax was in active development during the course of calibration, necessitating the ability to diagnose and flag potential conflicts in the model as part of the calibration process.
- Model was to be run on an HPC cluster, requiring a hands-off (but still robust) means of calibration that required little to no intervention.

Problem Statement

Can we create a means of calibration that is:

- Fast and robust;
- Amenable to use by epidemiologists with minimal background in statistics;
- Capable of identifying data/model conflicts and unwanted behaviours in model updates;
- "Hands-off" enough to be run on a cluster but informative enough for meaningful final analysis.

Problem Statement

Can we create a means of calibration that is:

- Fast and robust;
- Amenable to use by epidemiologists with minimal background in statistics;
- Capable of identifying data/model conflicts and unwanted behaviours in model updates;
- "Hands-off" enough to be run on a cluster but informative enough for meaningful final analysis.

Yes!

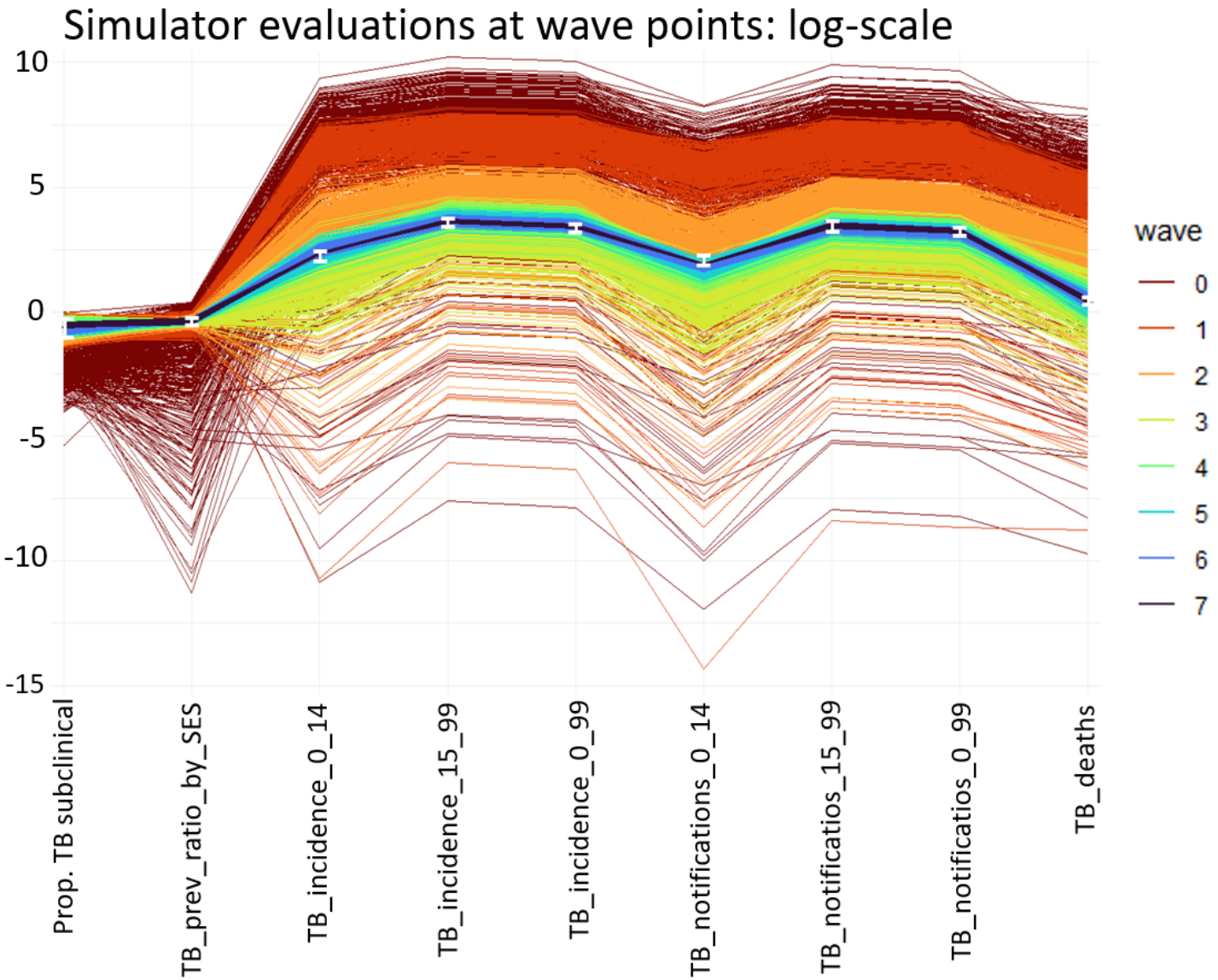
The hmer Package

I created an R package (`cran.r-project.org/package=hmer`) to allow

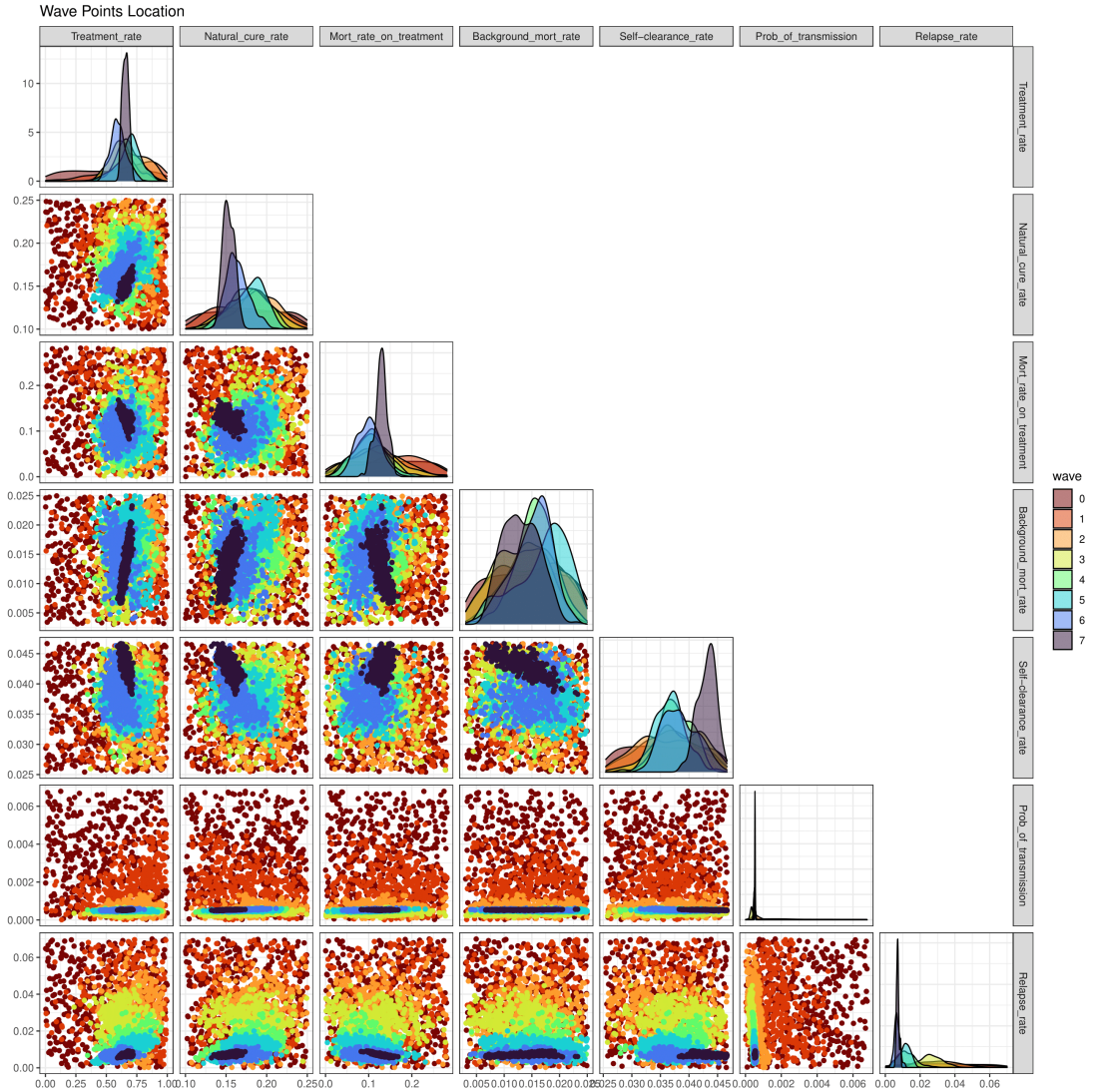
- Careful prior specifications to be determined and emulators to be trained
- Diagnostics to be performed to assess suitability
- Appropriate choices of implausibility measure and design for further waves to be made.

The application required the development of new techniques and methodology for proposing points from high dimensional spaces, automated identification of problematic combinations of inputs/outputs, emulator clustering based on active variables, and more.

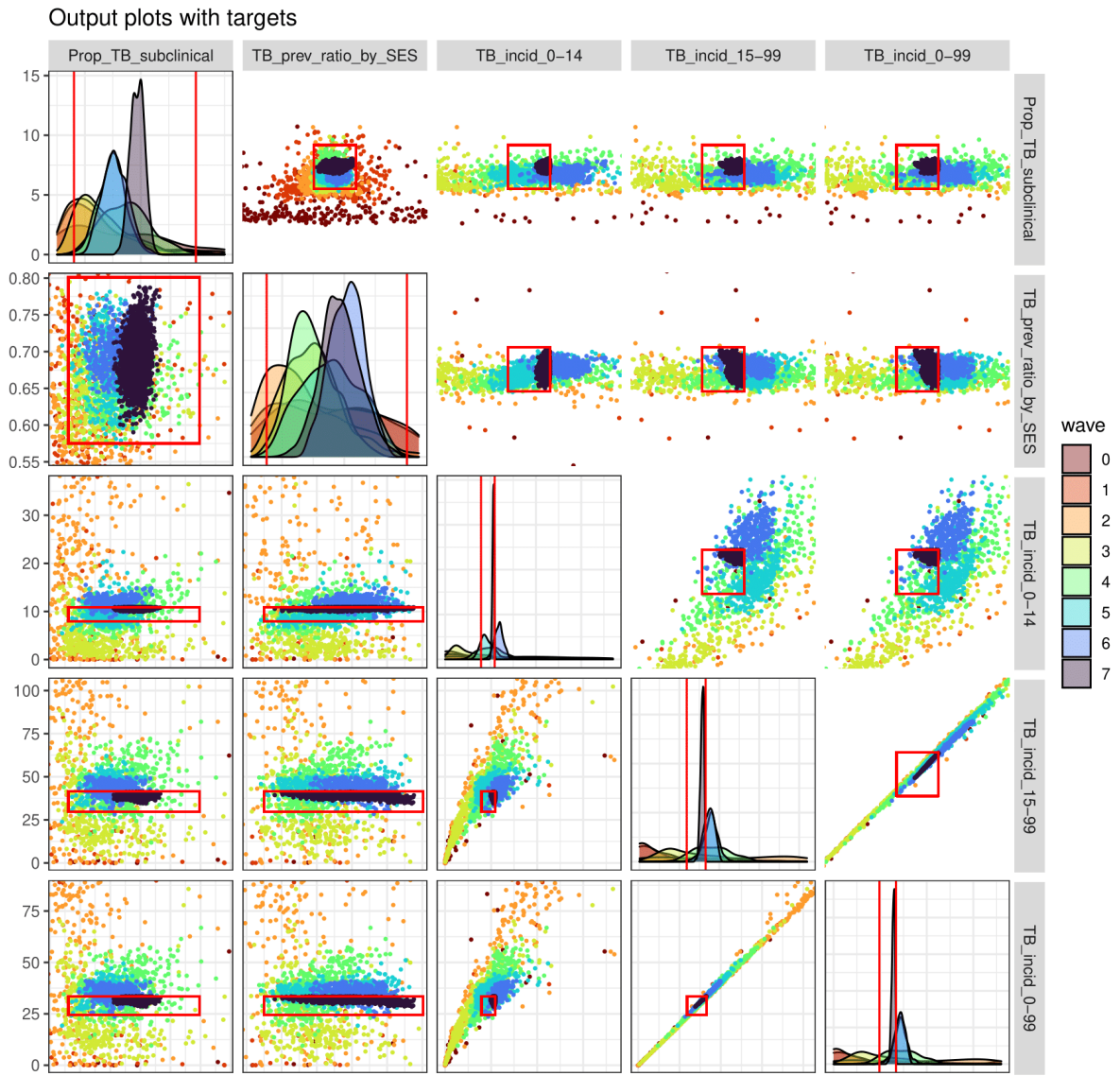
Results: Calibrated Countries



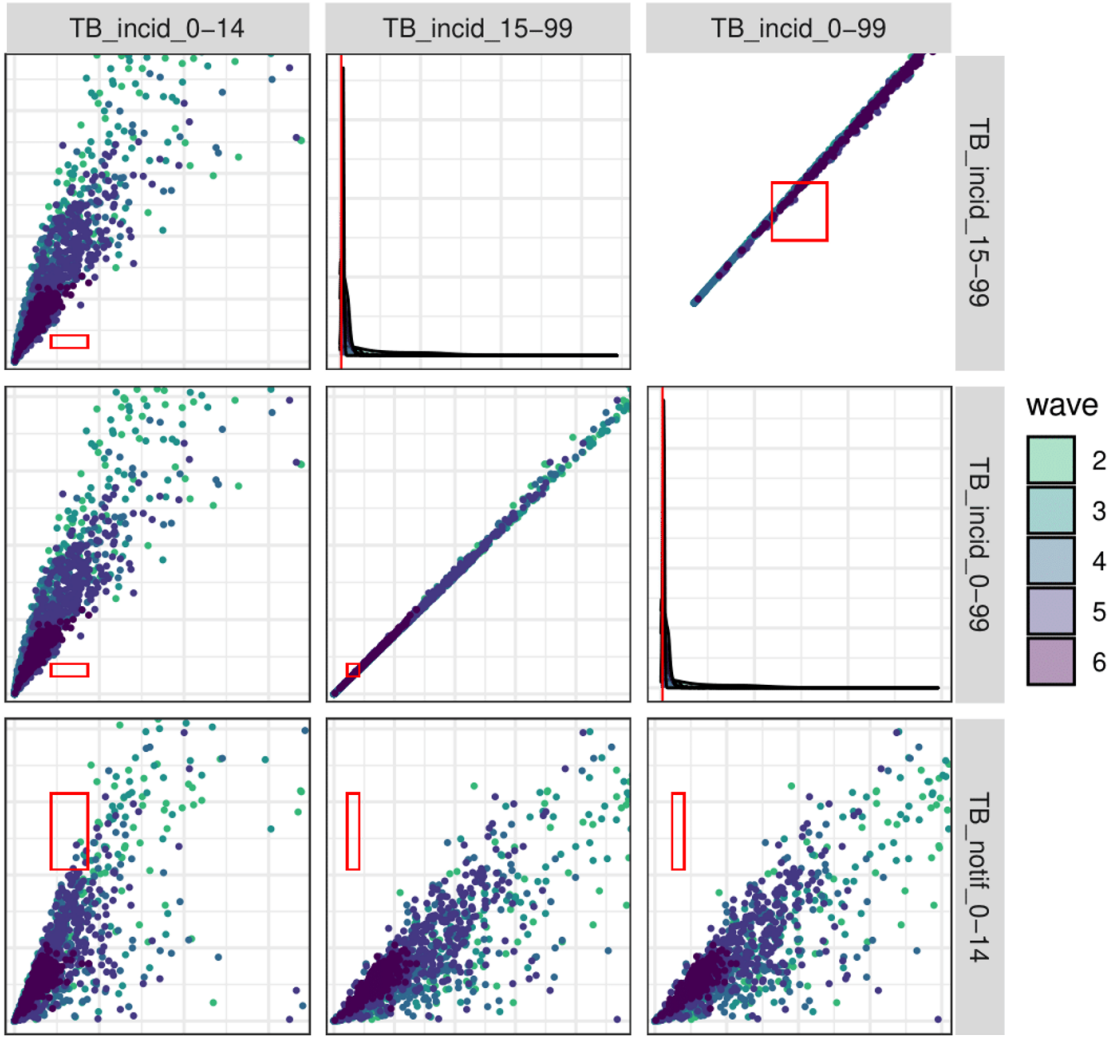
Results: Calibrated Countries



Results: Calibrated Countries



Results: Uncalibrated Countries



Adjoint Models and Derivative Emulation

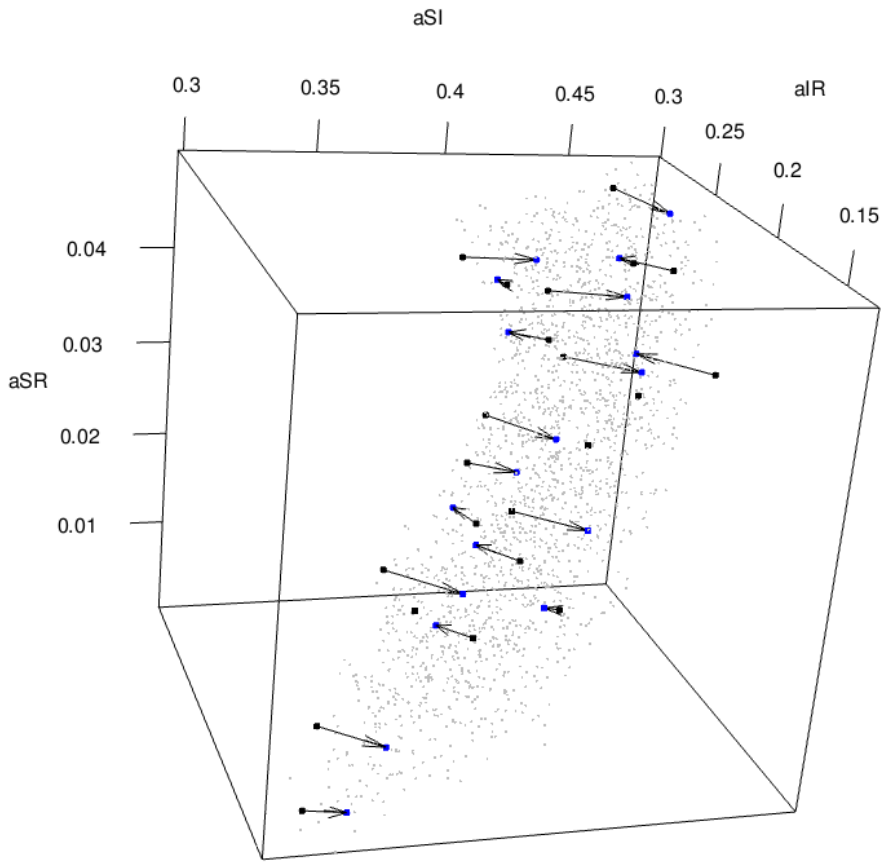
How sure can we be that we cannot find an acceptable match to data for a simulated output? If we had an *adjoint* model, we could consider derivatives, but not always possible!

Emulators can have a differentiable prior structure; the update formulae respect differentiation. Given data D, we can create a derivative emulator **without requiring derivative information from the simulator**.

$$\mathrm{E}_D[\partial_i g(x)] = \partial_i \mathrm{E}[g(x)] + \partial_i \mathrm{Cov}[g(x), D]\mathrm{Var}[D]^{-1}(D - \mathrm{E}[D]).$$

We can therefore predict both the simulator output and the corresponding derivative, with the appropriate uncertainty.

Simple Derivative Example



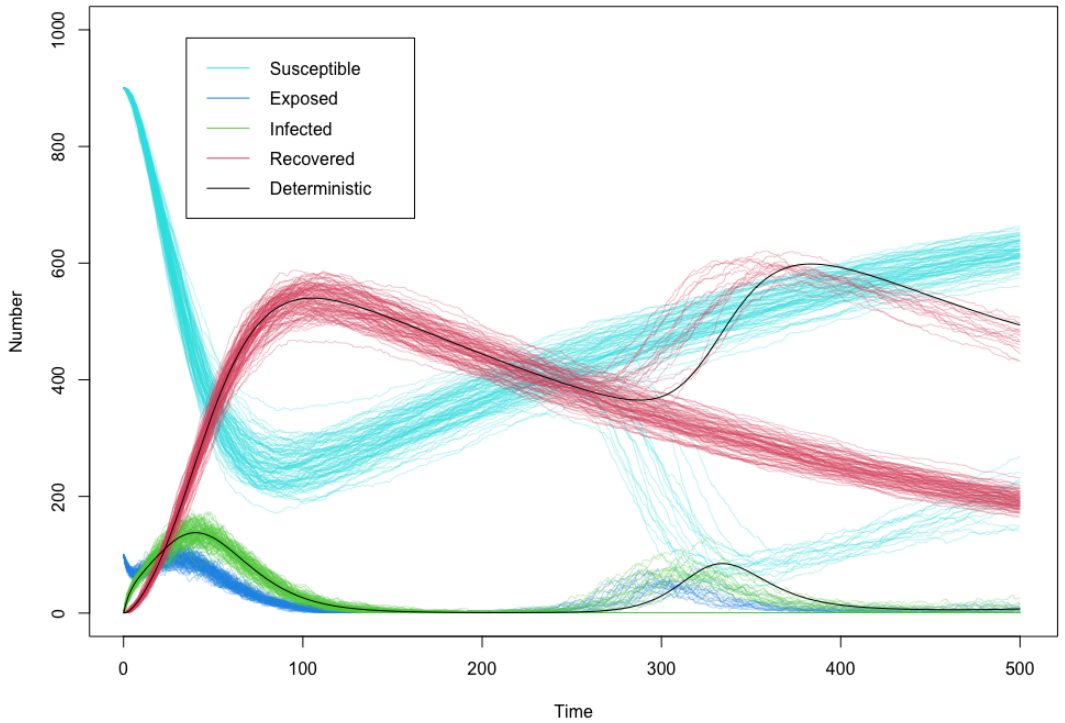
Either find a direction that should improve the proposed point, or show that there is no gradient direction in which to move.
Used in TBvax to strengthen the claim that countries were not likely to match to observed data.

TBvax: Summary

- The TBvax model has been used to inform WHO vaccine policy across the world as a result of the HME framework;
- Emulation allowed for fast model debugging and analysis during code development otherwise unavailable;
- 105 countries were matched to data, and the full space of allowed parameter space could be analysed;
- 10 countries were unable to be matched – visualisation and analysis of derivatives provided evidence for conflicts between observation and simulator;
- TBvax work has received additional funding from the WHO.

Stochastic Models

TBvax is a deterministic model; repeated evaluations at a chosen set of parameters gives the same simulator output. Many disease models are *stochastic*, introducing a new source of uncertainty.



Variance Emulation

One approach: consider realisations at a point x as *exchangeable*:

$$f_k(x) = \mathcal{M}(f(x)) + \mathcal{R}_k(f(x)),$$
$$[\mathcal{R}_k(f(x))]^2 \equiv V_k(x) = \mathcal{M}(V(x)) + \mathcal{R}_k(V(x)).$$

Then $\mathrm{Var}[\mathcal{R}_k(f(x))] = \mathrm{E}[\mathcal{M}(V(x))]$ under the assumption of second-order exchangeability.

We can construct an emulator for the *variance* across parameter space to provide an informed prior for the uncertainty on the emulator of the *output*. This requires prior statements about the structure of fourth-order quantities.

Multistate Behaviour

Multiple states of a model are most often seen in stochastic epidemiological simulators as 'take-off' vs 'die-out' of a disease. Even simple SIR models can exhibit this behaviour.

Unless handled appropriately, emulation will not be able to sufficiently capture the dynamics! We need a way to deal with this branching behaviour.

Our history matching question is also modified: we don't just care about if a parameter set x can match observed data, but **how often** its realisations match.

Multistate Emulation

The idea: for a collection of outputs, detect each state and perform emulation on each separately. It is important to ensure that we maximise the utility of the data we have; unlikely that the multistate behaviour is the same across parameter space.

We can also emulate the proportion of realisations in each state - leads to the possibility of interesting composite implausibility to favour points more likely to produce the observed reality.

Research Directions

Many interesting avenues, both in development of the theoretical underpinning and its specific application to the modelling community.

- Point proposal methods for multistate systems
- Covariance emulation
- Hierarchical model/meta-model emulation

Selected References I

Andrianakis, I. et al. (2015). Bayesian history matching of complex infectious disease models using emulation: A tutorial and a case study on HIV in uganda. *PLoS computational biology*, 11(1):e1003968.

Clark, R. A. et al. (2022). The impact of alternative delivery strategies for novel tuberculosis vaccines in low- and middle-income countries: A modelling study. *medRxiv*.

Craig, P. S. et al. (1998). Constructing partial prior specifications for models of complex physical systems. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1):37–53.

Cumming, J. A. and Goldstein, M. (2009). Small sample bayesian designs for complex high-dimensional models based on information gained using fast approximations. *Technometrics*, 51(4):377–388.

Goldstein, M. and Wooff, D. (2007). *Bayes Linear Statistics: Theory and Methods*, volume 716. John Wiley & Sons.

Iskauskas, A. et al. (2022). Emulation and history matching using the hmer package. *arXiv*.

Jackson, S. E. et al. (2020). Understanding hormonal crosstalk in arabidopsis root development via emulation and history matching. *Statistical Applications in Genetics and Molecular Biology*, 19(2).

Selected References II

Scarponi, D. et al. (2022). Demonstrating multi-country calibration of a tuberculosis model using new history matching and emulation package - hmer. *medRxiv*.

Vernon, I. et al. (2018). Bayesian uncertainty analysis for complex systems biology models: Emulation, global parameter searches and evaluation of gene functions. *BMC systems biology*, 12(1):1–29.

Vernon, I., Goldstein, M., and Bower, R. (2010). Galaxy formation: A bayesian uncertainty analysis. *Bayesian analysis*, 5(4):619–669.

Wilkinson, D. J. and Goldstein, M. (1995). Bayes linear covariance matrix adjustment for multivariate dynamic linear models.

Williamson, D., Goldstein, M., and Blaker, A. (2012). Fast linked analyses for scenario-based hierarchies. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(5):665–691.